Question 1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.508	0.038	-13.417	0.000
n_thorax	2.688	0.228	11.804	0.000
activityone	0.055	0.053	1.036	0.302
activitylow	-0.116	0.053	-2.184	0.031
activitymany	0.082	0.054	1.524	0.130
activityhigh	-0.415	0.054	-7.687	0.000

Figure 1: Summary of Model in Q1

	2.5 %	97.5 %
(Intercept)	0.560	0.649
n_thorax	9.311	23.127
activityone	0.952	1.173
activitylow	0.802	0.988
activitymany	0.977	1.208
activityhigh	0.595	0.734

Figure 2: Confidence Intervel in Q1

To determine the effect of thorax length and sexual activity on lifetime of male fruit fly. I fitted a gamma regression model with lifetime being the response variable, while thorax length and activity being the predictor variables. I centered and re-scaled the variables before fitting the model.

According to Figure 1. There is statistically significant evidence to show that thorax length has a positive association with expected lifetime (p-value is approximately zero). According to Figure 2, holding Activity constant, we are

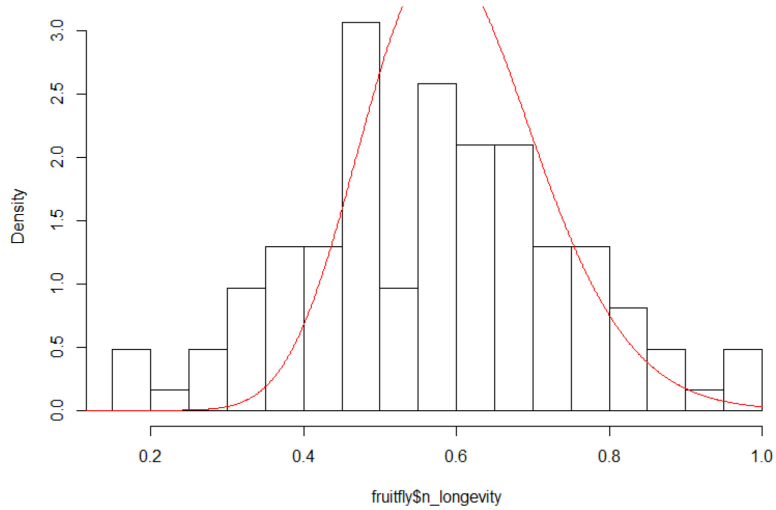

Figure 3: Histogram and Fitted Line

95% confident that expected lifetime increase between a factor of 9.311 and 23.127 if thorax length increases by 1. Since the 95% confidence interval of "One" and "Many" includes 1. I conclude there is no statistically significant evidence to show there is a difference in expected lifetime for male fruit flies among those kept isolated and those kept with 1 or 8 pregnant females, while holding thorax length constant. In the mean time, there is statistically significant difference in expected lifetime for male fruit flies among those kept isolated and those kept with 1 or 8 virgin females. Holding thorax length constant, We are 95% confident that expected lifetime decrease by a factor between 0.802 and 0.988 if kept with a virgin female instead of isolated, also, We are 95% confident that expected lifetime decrease by a factor between 0.595 and 0.734 if kept with eight virgin female instead of isolated.

As shown in Figure 3. The line fits the model fine. As shown in Figure 4,

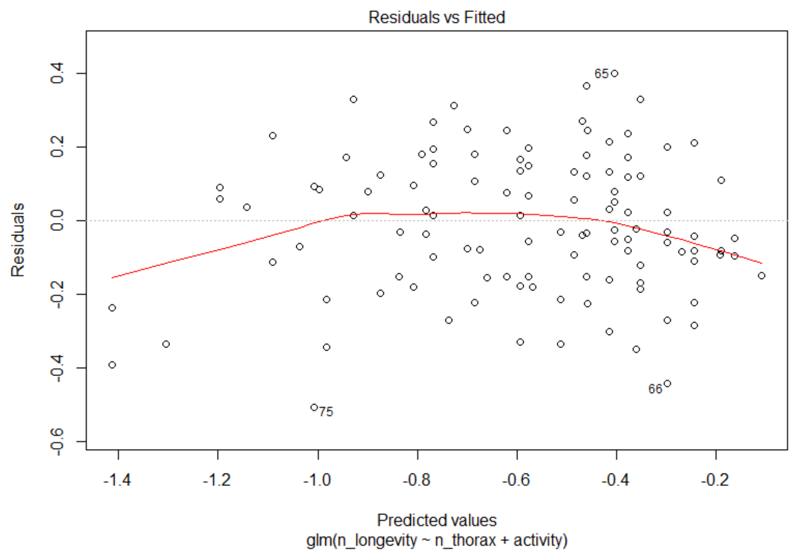

Figure 4: Residual vs Fitted

there is no distinctive pattern. Overall I conclude Gamma is an adequate fit.

It's interesting to see how sexual activity of male fruit flies plays a role on their expected life time. We already know males with longer thorax have longer lifespan. As my research on a total of 100 male fruit-flies has shown, males have little sexual activity tends to live longer compared with those that have sexual activity. In particular, those with low sexual activity lives longer than those with high sexual activity. Overall, higher level of sexual activity for male fruit flies lead to shorter lifespan.

Appendix

```r
install.packages("faraway")
data('fruitfly', package='faraway')
summary(fruitfly)
fruitfly$n_thorax = fruitfly$thorax - mean(fruitfly$thorax)
fruitfly$n_longevity = fruitfly$longevity / 100
model = glm(n_longevity ~ n_thorax + activity, family=Gamma(link='log'),
            data=fruitfly)
knitr::kable(summary(model)$coef, digits = 3)
shape = 1/summary(model)$dispersion
scale = exp(model$coef["(Intercept)"])/shape
hist(fruitfly$n_longevity, prob = TRUE, main = "", breaks = 25)
xSeq = seq(0, 1, len = 1000)
lines(xSeq, dgamma(xSeq, shape = shape, scale = scale), col = "red")
plot(model)
knitr::kable(exp(coef(model)), digits = 3)
knitr::kable(exp(confint(model)), digits = 3)
```

Figure 5: R Code for Question 1

Q2

Smoking Problem among youth is a major health concern. Research has shown regular use of chewing tobacco, snuff or dip is more common amongst children of White race compared with Hispanic and African American. While The proportion of having used a hookah or water-pipe at least once is the same between boys and girls with the same age and race.

The first null hypothesis is there is a no difference in the proportion of child chewing tobacco, snuff or dip at least once in the past 30 day between White American and African American or Hispanic-American. The second null hypothesis is there is no difference in the proportion of child having used a hookah or water-pipe on at least once between males and females with the same age and race.

For the first case, I choose to work with binomial distribution, since the dependent variable (chewing_tobacco_snuff_or) is discrete. Then I ran a dummy variable regression since the predictor variables (Race and Rural) are discrete. For the first hypothesis, I included an interaction term between Rural and Race, since Race has an effect on whether or not the child lives in rural area. For the second case, i choose binomial distribution and dummy variable regression for similar reasons. The independent variable is ever_tobacco_hookah_or_wa, while the predictor variables are Sex, Age and Race.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-3.657	0.096	-38.042	0.000
RuralUrbanRural	1.166	0.109	10.701	0.000
Raceblack	-1.006	0.256	-3.936	0.000
Racehispanic	-0.367	0.159	-2.314	0.021
Raceasian	-1.573	0.510	-3.083	0.002
Racenative	0.318	0.465	0.683	0.495
Racepacific	1.092	0.607	1.800	0.072
RuralUrbanRural:Raceblack	-0.706	0.341	-2.070	0.038
RuralUrbanRural:Racehispanic	-0.656	0.208	-3.149	0.002
RuralUrbanRural:Raceasian	0.346	0.684	0.506	0.613
RuralUrbanRural:Racenative	-0.694	0.570	-1.218	0.223
RuralUrbanRural:Racepacific	0.013	0.726	0.018	0.986

Figure 6: Summary of First Model

	2.5 %	97.5 %
(Intercept)	0.021	0.031
RuralUrbanRural	2.601	3.988
Raceblack	0.214	0.588
Racehispanic	0.505	0.942

Figure 7: Confidence Interval of Exp(Coefficients in Model 1)

As shown in Figure 6, we reject the first null hypothesis. We have statistically significant evidence at alpha = 0.001 that there is a difference in the proportion between White and African Americans. According to Figure 7. We are 95% confident that the proportion decreases by a factor between 0.214 to 0.588 in African Americans compared with White American. Also, we have statistically significant evidence at alpha = 0.05 that there is a difference in the proportion between White and Hispanic Americans. We are 95% confident that the proportion decreases by a factor between 0.505 to 0.942 in Hispanic Americans compared with White American. As shown in Figure 8, we cannot

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-8.145	0.185	-44.055	0.000
Age	0.414	0.011	36.078	0.000
Raceblack	-0.583	0.070	-8.317	0.000
Racehispanic	0.415	0.048	8.708	0.000
Raceasian	-0.510	0.117	-4.367	0.000
Racenative	0.143	0.190	0.749	0.454
Racepacific	0.997	0.268	3.724	0.000
SexF	0.043	0.043	1.011	0.312

Figure 8: Summary of Second Model

reject the second null hypothesis. We do not have statistically significant evidence that there is a difference in the proportion between males and females.

Appendix

```r
load(file = "smoke.Rdata")
summary(smoke)
smokeSub = smoke[smoke$Age >= 10, ]
#model for part 1:
model = glm(chewing_tobacco_snuff_or ~ RuralUrban * Race,
            family=binomial, data=smokeSub)
knitr::kable(summary(model)$coef, digits = 3)
knitr::kable(exp(confint(model)), digits = 3)

#model for part 2:
model3 = glm(ever_tobacco_hookah_or_wa ~ Age + Race + Sex,
             family=binomial, data=smokeSub)
summary(model3)
knitr::kable(summary(model3)$coef, digits = 3)
```

Figure 9: R Code for Question 2